

Using the goodenuffR package

Journalism 303: everything...in its right place

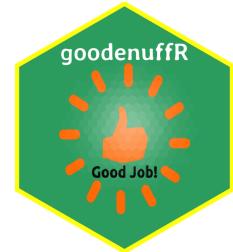
Martin Frigaard

2021-10-14

goodenuffR = good enuff practices in R computing



Outline



- 1) The Problem
- 2) A Solution
- 3) Installation
- 4) Quick Start

5) Package functions

- `goodenuff_code()`
- `goodenuff_collab()`
- `goodenuff_data()`
- `goodenuff_doc()`
- `goodenuff_results()`
- `goodenuff_files()`

6) Using **goodenuffR**

The Problem



| *Where is my stuff?*

THE VERGE: FILE NOT FOUND

THE VERGE

REPORT

FILE NOT FOUND

A generation that grew up with Google is forcing professors to rethink their lesson plans

By Monica Chin | @mcsquared96 | Sep 22, 2021, 8:00am EDT
Illustrations by Micha Huijen

The Problem

Where is my stuff?



12:15

Emily Riederer and 4 others liked

Jenny Bryan @JennyBryan · 40m

Knowing how to store stuff in a particular place, with a intentional name, is SO IMPORTANT

dan nguyen @dancow · 2h

omg turns out I wasn't completely bullshitting all the times when I fussed to students that "file not found" would be their most common and soul-destroying "bug" when trying to learn programming

Show this thread

Chelsea Parlett-Pelleriti and 3 others liked

Matthew Kay @mjskay · 42m

It's September, time to talk about how the kids don't know what directories are

Chelsea Parlett-Pelleriti and 3 others liked

Matthew Kay @mjskay · 42m

It's September, time to talk about how the kids don't know what directories are

Chelsea Parlett-Pelleriti and 3 others liked

Matthew Kay @mjskay · 42m

It's September, time to talk about how the kids don't know what directories are

A Solution



Based on the excellent paper, *Good enough practices in scientific computing*

by Greg Wilson, Jennifer Bryan, Karen Cranston, Justin Kitzes, Lex Nederbragt, Tracy K. Teal

PLOS COMPUTATIONAL BIOLOGY

OPEN ACCESS

PERSPECTIVE

Good enough practices in scientific computing

Greg Wilson , Jennifer Bryan , Karen Cranston , Justin Kitzes , Lex Nederbragt , Tracy K. Teal

Published: June 22, 2017 • <https://doi.org/10.1371/journal.pcbi.1005510>

Article	Authors	Metrics	Comments	Media Coverage

A Solution



The paper focuses on...

"a minimum set of tools and techniques that we believe every researcher can and should consider adopting"

Specifically

- Data management
- Software
- Collaboration
- Project organization
- Tracking changes
- Manuscripts

A Solution

goodenuffR focuses on

- Data management
- Project organization

Specifically

```
-- CITATION
-- README
-- LICENSE
-- requirements.txt
-- data
|   |-- birds_count_table.csv
-- doc
|   |-- notebook.md
|   |-- manuscript.md
|   |-- changelog.txt
-- results
|   |-- summarized_results.csv
-- src
|   |-- sightings_analysis.py
|   |-- runall.py
```



Installation



Available on GitHub here:

<https://github.com/mjfrigaard/goodenuffR>

Install **goodenuffR** using the code below:

```
install.packages("devtools")
devtools::install_github("mjfrigaard/goodenuffR")
```

Quick Start

In a fresh RStudio Session



The screenshot shows a fresh RStudio session with the following interface elements:

- Environment pane:** Shows tabs for Environment, History, Connections, and Tutorial. The Global Environment section is expanded, showing a message "Environment is empty".
- Console pane:** Displays the R startup message for version 4.1.1, followed by a welcome message: "R is free software and comes with ABSOLUTELY NO WARRANTY. You are welcome to redistribute it under certain conditions. Type 'license()' or 'licence()' for distribution details." It also mentions Natural language support and R's collaborative nature.
- Packages pane:** Shows the System Library with a list of packages and their details. The packages listed are: abind, acepack, anytime, AppliedPredictiveMo..., arrow, arsenal, ash, AsioHeaders, askpass, assertive, and assertive.base. Each package has a brief description, its version number, and two circular checkboxes for 'Install' and 'Update'.

Quick Start



1) Install `devtools`, then install and load `goodenuffR`

```
Console Terminal × Jobs ×
R 4.1.1 · ~/R/ ⓘ
> install.packages("devtools")
trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.1/devtools_2.4.2.tgz'
Content type 'application/x-gzip' length 393913 bytes (384 KB)
=====
downloaded 384 KB

The downloaded binary packages are in
  /var/folders/3p/wzkys03s6p1cvmn8yzm934400000gn/T//Rtmp4rq8qj downloaded_packages
> devtools::install_github("mjfrigaard/goodenuffR")
Using github PAT from envvar GITHUB_PAT
Downloading GitHub repo mjfrigaard/goodenuffR@HEAD
✓  checking for file '/private/var/folders/3p/wzkys03s6p1cvmn8yzm934400000gn/T/Rtmp4rq8qj/remotesf16118ee2b1d/mjfrigaard-goodenuffR-3da8f43/DESCRIPTION' ...
-  preparing 'goodenuffR':
✓  checking DESCRIPTION meta-information ...
-  checking for LF line-endings in source and make files and shell scripts
-  checking for empty or unneeded directories
-  building 'goodenuffR_0.1.0.tar.gz'

* installing *source* package 'goodenuffR' ...
** using staged installation
** R
** data
*** moving datasets to lazyload DB
** inst
** byte-compile and prepare package for lazy loading
** help
*** installing help indices
** building package indices
** testing if installed package can be loaded from temporary location
** testing if installed package can be loaded from final location
** testing if installed package keeps a record of temporary installation path
* DONE (goodenuffR)
> library(goodenuffR)
```

Quick Start



2) Enter a project name

3) Enter the folder path for your project

```
goodenuffR::goodenuff_project(  
  project_name = "my awesome project",  
  folder_path = "/Users/mjfrigaard/Projects")
```

TIP: use *getwd()* in the Console or *pwd* in the Terminal to locate current folder

Quick Start

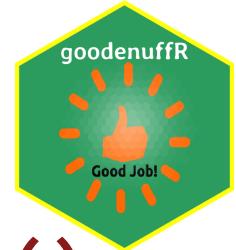


RStudio will open a new session for your project

The screenshot shows the RStudio interface with a new project named "my_awesome_project".

- Environment:** Shows the Global Environment pane with the message "Environment is empty".
- Console:** Displays the R startup message and basic information about the R version and platform.
- Files:** Shows the project directory structure:
 - Home > Projects > my_awesome_project
 - Contents:
 - .. (empty folder)
 - .Rproj.user (file)
 - my_awesome_project.Rproj (file, size 205 B, modified Oct 14, 2021, 7:47 AM)

Quick Start



4) Run `goodenuffR::goodenuff_files()`

```
Console | Terminal ✘ | Jobs ✘
R 4.1.1 · ~/Projects/my_awesome_project/ ↵

R version 4.1.1 (2021-08-10) -- "Kick Things"
Copyright (C) 2021 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> goodenuffR::goodenuff_files()
trying URL 'https://creativecommons.org/publicdomain/zero/1.0/legalcode.txt'
downloaded 7048 bytes

trying URL 'https://raw.githubusercontent.com/rstudio/rmarkdown/main/inst/rmarkdown/templates/git_
_document/skeleton/skeleton.Rmd'
Content type 'text/plain; charset=utf-8' length 691 bytes
=====
downloaded 691 bytes
```

Quick Start



Now you're good (enuff) to go!

A screenshot of the RStudio Project Explorer. The title bar shows "Files", "Plots", "Packages", "Help", and "Viewer". Below the title bar are buttons for "New Folder", "Delete", "Rename", and "More". The path "Home > Projects > my_awesome_project" is displayed. The main area lists project files and folders: ".Rproj.user", "my_awesome_project.Rproj", "CITATION", "LICENSE", "README.Rmd", "code", "data", "docs", "requirements.txt", and "results".

	Name
<input type="checkbox"/>	..
<input type="checkbox"/>	.Rproj.user
<input type="checkbox"/>	my_awesome_project.Rproj
<input type="checkbox"/>	CITATION
<input type="checkbox"/>	LICENSE
<input type="checkbox"/>	README.Rmd
<input type="checkbox"/>	code
<input type="checkbox"/>	data
<input type="checkbox"/>	docs
<input type="checkbox"/>	requirements.txt
<input type="checkbox"/>	results

Package functions



goodenuff_code()

Paper

- *readable, reusable, and testable*
- *short, single-purpose functions with clearly-defined inputs and outputs*
- *name all files to reflect their content or function*

goodenuffR

```
goodenuffR::goodenuff_code()  
# └── code  
#   ├── 01-import.R  
#   ├── 02-tidy.R  
#   ├── 03-wrangle.R  
#   ├── 04-visualize.R  
#   ├── 05-model.R  
#   ├── 06-communicate.R  
#   └── runall.R
```

Package functions



goodenuff_code()

Paper

- *Place a brief explanatory comment at the start of every program*

goodenuffR

Each `code` file has a header:

```
=====
# This is code to create:
# Authored by and feedback to:
# MIT License
# Version:
=====
```

Package functions

goodenuff_collab()



Paper

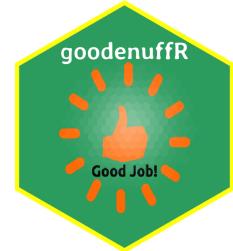
- *Make the project citable*
- *Make the license explicit*
- *Create an overview of your project*
- *Make dependencies and requirements explicit*

goodenuffR

```
goodenuffR::goodenuff_collab()  
# └─ CITATION  
# └─ LICENSE  
# └─ README.Rmd  
# └─ requirements.txt
```

Package functions

goodenuff_data()



Paper

- *Save the raw data*
- *Record all the steps used to process data*

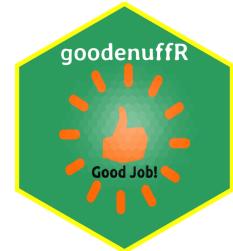
goodenuffR

```
goodenuffR::goodenuff_data()  
# data  
# └── README.md  
# └── raw
```

*Folder for keeping **raw**/ data separate from other data files*

*Document in **README.md***

Package functions



goodenuff_data() (BONUS!)

data/README.md provides some guidance on storing and documenting data.

Project Data

For guidance please see:

<https://help.osf.io/hc/en-us/articles/360019739054-How-to-Make-a-Data-D->

Additional Resources

1. Sharing data

<http://bit.ly/data-4-sharing>

2. Data in spreadsheets

<http://bit.ly/data-in-sheets>

3. External data (<https://r-pkgs.org/>)

<https://r-pkgs.org/data.html>

source:

<https://swcarpentry.github.io/good-enough-practices-in-scientific-computing/>

Package functions



goodenuff_doc()

Paper

- Put text documents associated with the project in the `doc` directory
- Add a file called `CHANGELOG.txt` to the project's docs subfolder
- a running lab notebook describing various ideas for the project and how these were implemented

goodenuffR

```
goodenuffR::goodenuff_doc()  
# docs  
#   ├── changelog.txt  
#   ├── manuscript.Rmd  
#   └── notebook.Rmd
```

Package functions

goodenuff_results()

Paper

- *Put raw data and metadata in a `data/` directory and files generated during cleanup and analysis in a `results/` directory*

goodenuffR

```
goodenuffR::goodenuff_results()  
# results  
#   └── figures  
#   └── manuscript  
#   └── tables
```



Package functions



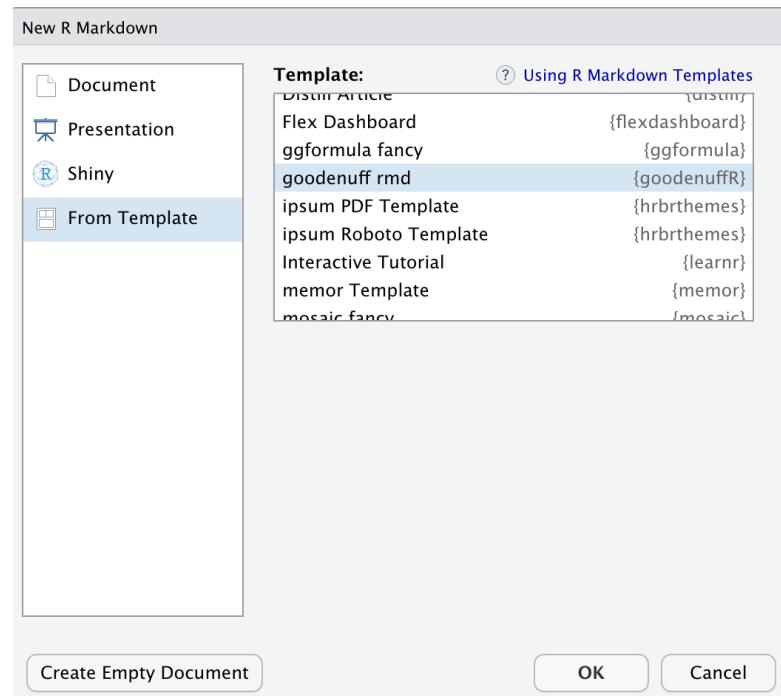
goodenuff_files()

```
goodenuffR::goodenuff_files()
├── CITATION
├── LICENSE
├── README.Rmd
└── code
    ├── 01-import.R
    ├── 02-tidy.R
    ├── 03-wrangle.R
    ├── 04-visualize.R
    ├── 05-model.R
    ├── 06-communicate.R
    └── runall.R
├── data
│   ├── README.md
│   └── raw
└── docs
    ├── changelog.txt
    ├── manuscript.Rmd
    └── notebook.Rmd
├── my_project.Rproj
└── requirements.txt
└── results
    ├── figures
    ├── manuscript
    └── tables
```

BONUS: RMarkdown Template



goodenuffR also comes with a `rmarkdown` template:



Thank you!

Feedback to

- **@mjfrigaard** on Github and Twitter
- **mjfrigaard(at)pm.me** email

